

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE CONTENT

A Seven day Internship Program on “Big data and Hadoop”

An internship Program on “Big data and Hadoop” organized by Computer science and engineering department and presented by Mr. Pankaj Singh and Mr. Ravi Kumar, Revert Technologies.

The seven day internship Program aimed at learning with hands on session on project development on the various platforms of big data and Hadoop. The concentration was on Map and reduce, pig, hive, Hbase, flume, scoop and zookeeper.

The workshop comprised of the following themes:

Day 1

A. Introduction to Big Data and Hadoop

- What is a Data?
- Type of Data
- Need of Big Data
- Characteristics of Big Data

B. Different Components of Hadoop

C. Big Data Technology

- Traditional IT approach
- Big Data Capabilities
- Milestones of Hadoop

Day 2

A. Software Introduction

- VMware Player
- VMware installed with BIOS system
- Horton Works Sand Box Introduction

B. Hadoop Architecture

- Hadoop cluster
- Hadoop Core Services

- Hadoop Core Components
- Map reduce Introduction
- HDFS

Day 3

A. Starting With Hadoop

- Map reduce Analogy
- Map reduce Example
- Map Execution
- Real time Using With Hadoop

B. Pig

- Introduction to Apache Pig
- Components of pig
- How to Works pig, or Data model
- Pig vs. SQL
- Pig Execution Modes

Day 4

A. Pig

- Map Reduce vs. Apache Pig
- Different Data Types in Pig
- Modes of Execution in Pig
- Local Mode
- Execution Mechanism
- Grunt Shell
- Scrip
- Pig Commands

- Examples Of pig
- 6.1 Word Count
- 6.2 Batting Examples

Day 5

A. Hive

- Hive Introduction
- Hive characteristics
- System Architecture & components of Hive
- Query Compiler
- Hive Server
- HIVE Architecture
- HIVE Meta Store
- HIVE Integration with Hadoop
- HIVE Query Language (Hive QL)
- SQL Vs Hive QL

B. HBase

- Introduction of HBase
- Characteristics of HBase
- HBase Architecture
- HBase Vs RDBMS
- HBase Shell Commands

Day 6

A. HBase Architecture

- Clients
- REST
- Thrift

Java Based

- AVTO

B. MapReduce Integration

- MapReduce over HBase
- HBase Admin
- Schema Definition
- MapReduce OR Distributed Mode

Day 7

A. SQOOP


- Introduction to SQOOP
- MySQL Client and Server Installation
- How to Connect to Relational Database
- Using SQOOP
- Different SQOOP Commands
- Different Flavors of Import
- Export
- HIVE Imports

B. Introduction Of Zookeeper

- Features of Zookeeper
- Use of ZooKeeper
- Zookeeper Data Model

By end of this workshop the students will be able to Implement a Hadoop Project. Since Bigdata & Hadoop is the recent technology and which is not included in curriculum, the workshop was conducted to fulfill the gap and we have attained the PO's **Engineering Knowledge, Problem Analysis, Design/Development of solutions, Modern Tool Usage, Life Long Learning**. This Workshop on Bigdata & Hadoop has provided an ability to understand analyze and develop computer programs in the areas related to algorithms, multimedia, web design and networking for computer based systems of varying complexity.

The practical sessions have helped the students to find solutions for real world problems with a broad range of programming languages and open source.


PRINCIPAL SIGNATURE WITH SEAL
Dr. R. S. Shrinani
Principal
MVJ College of Engineering
Near ITPS, Channarayana
Bangalore-560 067

